Distributed data fusion in the Dempster-Shafer framework
Orakanya Kanjanatarakul, Thierry Denoeux

To cite this version:
Orakanya Kanjanatarakul, Thierry Denoeux. Distributed data fusion in the Dempster-Shafer framework. 2017 12th System of Systems Engineering Conference (SoSE), Jun 2017, Waikoloa, United States. pp.1-6, 10.1109/SYSOSE.2017.7994954 . hal-03781559

HAL Id: hal-03781559
https://hal.science/hal-03781559
Submitted on 20 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Distributed Data Fusion in the Dempster-Shafer framework

Orakanya Kanjanatarakul
Chiang Mai Rajabhat University
Chiang Mai, Thailand
Email: orakanyaa@gmail.com

Thierry Denœux
Sorbonne Universités
Université de Technologie de Compiègne, CNRS,
UMR 7253 Heudiasyc, Compiègne, France
Email: tdenoeux@utc.fr

Abstract—Dempster-Shafer theory is a formal framework for reasoning and decision-making under uncertainty. A cornerstone in this formalism is Dempster’s rule, which provides a mechanism for combining belief functions representing independent pieces of evidence. This rule has been used extensively in information fusion applications. In this paper, we consider the situation where several agents are located at the nodes of a network and can communicate only with their neighbors. We show that synchronous or asynchronous linear consensus mechanisms can be used to combine the agents’ belief functions in a distributed way. After convergence, a consensus state is reached, in which all agents hold the same belief, which is equal to the orthogonal sum of the agents’ initial belief functions. Simulation results with a simple distributed classification are reported.

I. INTRODUCTION

Dempster-Shafer (DS) theory [1], [2] is a well-known and widely used framework for uncertain reasoning and information fusion. In the past thirty years, it has been used extensively in a large number of applications including classification [3], [4], clustering [5], [6], image segmentation [7], scene perception [8], etc. DS theory has two main components: at the static level, pieces of evidence about some question of interest are represented by belief functions (i.e., completely monotone set functions); at the dynamic level, independent items of evidence are pooled using the so-called Dempster’s rule of combination (or orthogonal sum). Although alternative rules have been proposed, especially to combine highly conflicting [9] or dependent pieces of evidence [10], Dempster’s rule remains the cornerstone of DS theory.

To our knowledge, all implementations of Dempster’s rule have assumed centralized fusion scheme in which a single agent receives pieces of evidence from several sources and combines them. In this paper, we consider a distributed scheme whereby several agents independently collect evidence and exchange information via a static or dynamic communication network. This application is motivated by a System of Systems composed of a fleet of vehicles equipped with sensor and communicating through an ad hoc network (see, e.g. [11], [12]). The overall objective of this application is to enhance each vehicle’s perception and situation awareness of a complex dynamic traffic scene through the multiplicity of sensors and the communication capabilities of the agents. To process vast amounts of data collected at the nodes of ad hoc networks, centralized or hierarchical fusion schemes can no longer be implemented. We need to develop distributed fusion algorithms, which allow the different agents to exchange information locally with their neighbors in the network and to construct, collectively, a shared representation of the environment. This objective requires the availability of algorithms for performing various combination and reasoning tasks (including object association [13], tracking and classification) in a distributed way.

In this paper, we focus on a specific subtask, which is the distributed combination of belief functions by Dempster’s rule. We assume that the agents have already agreed on a common frame of discernment, or universal set, on which the belief functions to be combined have been defined. We show that average consensus algorithms make it possible to combine the agents’ belief functions in a distributed way, resulting in a consensus state in which all agents hold the same belief function, which is the orthogonal sum of all initial individual belief functions.

The rest of this paper is organized as follows. Some background on DS theory is first recalled in Section II. The distributed implementations of Dempster’s rule are then described in Section III. Simulation results are then reported in Section IV, and Section V concludes the paper.

II. DEMPSTER-SHAFER THEORY

Let \( \Omega \) be a finite set of possible answers to some question, one and only one of each is true. A mass function on \( \Omega \) is a mapping \( m : 2^\Omega \to [0,1] \) such that

\[
\sum_{A \subseteq \Omega} m(A) = 1.
\]

It is normalized if \( m(\emptyset) = 0 \). In DS theory, a mass function represents a piece of evidence, and \( m(A) \) represents the probability that the evidence tells us that the truth lies in \( A \), and nothing more specific. A belief function \( Bel : 2^\Omega \to [0,1] \) can be computed from a normalized mass function \( m \) using the following formula, \( Bel(A) = \sum_{B \subseteq A} m(B) \). The quantity \( Bel(A) \) represents a total degree of support given to the proposition that the truth lies in \( A \). A related notion is the plausibility function, defined as

\[
Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) = 1 - Bel(\overline{A}),
\]
where $\overline{A}$. The quantity $P(A)$ reflects the lack of support given to the proposition that the truth does not lie in $A$.

Two mass functions $m_1$ and $m_2$ on $\Omega$ representing independent items have evidence can be combined by Dempster’s rule. The combined mass function $m_1 \oplus m_2$, called the orthogonal sum of $m_1$ and $m_2$, is defined as follows,

$$(m_1 \oplus m_2)(A) = K \sum_{B \subseteq C = A} m_1(B)m_2(C)$$

for all nonempty subset $A$ of $\Omega$, where

$$K = \left[ \sum_{B \subseteq C \neq \emptyset} m_1(B)m_2(C) \right]^{-1}$$

is a normalizing constant, and $(m_1 \oplus m_2)(\emptyset) = 0$. Dempster’s rule can also be computed using the commonality function. The commonality function $Q : 2^{\Omega} \to [0, 1]$ associated to a mass function $m$ is defined by

$$Q(A) = \sum_{B \supseteq A} m(B),$$

for all $A \subseteq \Omega$. Conversely, $m$ can be recovered from $Q$ using the following formula,

$$m(A) = \sum_{B \supseteq A} (-1)^{|B| - |A|} Q(B)$$

for all $A \subseteq \Omega$. If $Q_1$ and $Q_2$ are the commonality functions associated with two mass functions $m_1$ and $m_2$, then the commonality function $Q_1 \oplus Q_2$ associated with $m_1 \oplus m_2$ is $Q_1 \oplus Q_2 = KQ_1Q_2$, where $K$ is the normalizing constant (2). Dempster’s rule is commutative and associative. To combine $n$ mass functions $m_1, \ldots, m_n$, we may proceed as follows:

1) Compute the commonality functions $Q_1, \ldots, Q_n$ using (3);
2) Compute the product $Q = Q_1 \ldots Q_n$;
3) Compute the unnormalized mass function $m$ from $Q$ using (4);
4) Compute the corresponding normalized mass function $m^*$ such that $m^*(A) = Km(A)$ for all $A \neq \emptyset$ with $K = 1/(1 - m(\emptyset))$ and $m^*(\emptyset) = 0$.

III. DISTRIBUTED COMBINATION

In this section, we will show how Dempster’s rule can be implemented in a distributed way. Let us assume that $n$ agents are located at the nodes of an undirected communication graph $\mathcal{G}(t) = (\mathcal{N}, \mathcal{E}(t))$, where $\mathcal{N} = \{1, \ldots, n\}$ is the set of nodes, $\mathcal{E}(t)$ is the set of edges and $t$ is a discrete time index. Each agent $i$ holds a mass function $m_i$ and can communicate only to its neighbors in the graph. We wish to design distributed procedures whereby each agent can combine its mass function with those of other agents by Dempster’s rule. As a result, a consensus will be reached, each agent having the same mass function $m^* = m_1 \oplus \ldots \oplus m_n$. The key idea is to turn the problem of combining mass functions by Dempster’s rule into one of averaging certain quantities, and then to use linear average consensus algorithms [14].

In Section III-A, we will assume the graph to be constant, $\mathcal{G}(t) = \mathcal{G}$ and the consensus strategy to be synchronized. In Section III-B, we will consider the more general situation of a time-varying graph and an asynchronous updating mechanism.

A. Static consensus

Let us first show that Dempster’s rule can be computed by averaging certain quantities. With the notations of Section II, we have

$$\log Q(A) = \sum_{i=1}^{n} \log Q_i(A) = n \left( \frac{1}{n} \sum_{i=1}^{n} \log Q_i(A) \right)$$

for any $A \subseteq \Omega$. Consequently, the orthogonal sum $m^* = m_1 \oplus \ldots \oplus m_n$ can be computed by averaging the terms log $Q_i(A)$ for all subsets $A$ of $\Omega$, multiplying by $n$, converting the combined commonality function $Q$ into an unnormalized mass function $m$ using (4), and renormalizing.

Let $A_1, \ldots, A_N$ be the nonempty subsets of $\Omega$ (propositions) arranged in some order. Let $x_{ik}(0)$ denote the log-commonality log $Q_i(A_k)$ of proposition $A_k$ for agent $i$, $x_i(0)$ the $N$-vector $(x_{i1}, \ldots, x_{iN})$ of log-commonalities for agent $i$, $x_k(0)$ the $n$-vector $(x_{1k}, \ldots, x_{nk})$ of log-commonalities of $A_k$ for the $n$ agents, and $x(0)$ the initial $nN$-dimensional state vector $(x_{11}(0), \ldots, x_{N1}(0))$ composed of the complete initial log-commonality functions for the $n$ agents. Let $\alpha_k$ denote the average log-commonality of proposition $A_k$,

$$\alpha_k = \frac{1}{n} \sum_{i=1}^{n} x_{ik}(0),$$

and $\alpha$ the $N$-vector $\alpha = (\alpha_1, \ldots, \alpha_N)$.

In this section, we assume the communication graph $\mathcal{G}$ to be constant and strongly connected, i.e., there exists a path between any two nodes in the graph. We further assume that the partial state vectors $x_{ik}(t)$ are updated at time $t$ using linear equations

$$x_{ik}(t+1) = Wx_{ik}(t), \quad k = 1, \ldots, N$$

where $W$ is an $n \times n$ square matrix. Let $W$ be the $nN \times nN$ matrix $W = I_N \otimes W$, where $I_N$ is the identity matrix of size $N$ and $\otimes$ is the Kronecker product. Then, the $N$ update equations (6) can be written more compactly as

$$x(t+1) = Wx(t).$$

According to a well-known result in the theory of Markov chains (see, e.g., Theorem 3.1 in [14]), if

1) Matrix $W$ is doubly stochastic, i.e., $w_{ij} \geq 0$ and $\sum_i w_{ij} = \sum_j w_{ij} = 1$ for all $(i, j)$;
2) $G_W = G$, where $G_W$ is the graph associated with $W$, with nodes $\mathcal{N}$ and edges $E_W = \{(i, j) | w_{ij} > 0\}$,

then matrix $W$ solves the average consensus problem, i.e., we have $\lim_{t \to \infty} x_{ik}(t) = \alpha_k$ and, consequently,

$$\lim_{t \to \infty} x(t) = \alpha.$$
A particular matrix verifying the conditions above is the matrix of Metropolis-Hastings weights [15],

\[ w_{ij} = \begin{cases} \frac{1}{\max(d(i),d(j)) + 1} & \text{if } (i, j) \in \mathcal{E} \text{ and } i \neq j \\ 1 - \sum_{j=1, j \neq i}^{n} w_{ij} & \text{if } i = j, \end{cases} \]  

(9)

where \( d(i) \) is the number of neighbors of node \( i \). For this choice of matrix \( W \) (and the corresponding \( \mathbf{W} \)), the convergence property (8) holds. Consequently, the network converges to a consensus state where each node holds the same information \( \alpha \), from which the combined commonality function \( Q \) can be retrieved as \( Q(A_k) = \exp(n\alpha_k) \), and the combined mass function \( m^* \) can be computed using (4) and renormalization.

B. Dynamic consensus

In the approach described in Section III-A, the communication graph \( \mathcal{G} \) was assumed to be static, and it was assumed that nodes exchange information and update their states simultaneously. In this section, we relax these limitations by considering time-varying graphs and asynchronous communication between nodes. More specifically, we consider the symmetric gossip scheme [16], [14], in which at each time step a node \( i \) transmits its information to one of its neighbors \( j \), which in turn transmits back it information to \( i \). After this information exchange, both nodes update their state using a consensus scheme.

More formally, given a graph \( \mathcal{G} = (\mathcal{N}, \mathcal{E}) \), assume that at each time \( t \) a link \((i, j) \in \mathcal{E}\) is selected at random, and matrix \( \mathbf{W}(t) \) is defined as

\[ W(t) = W^{ij} = I_n - \frac{1}{2}(e_j - e_i)(e_j - e_i)^T, \]  

(10)

where \( I_n \) is the identity matrix of dimension \( n \), and \( e_i \) is the \( n \)-vector of all zeros except for the \( i \)-th entry which is set to one. All matrices \( W(t) \) are doubly stochastic and \( \mathcal{G}_W(t) \subseteq \mathcal{G} \). Furthermore, let \( \mathbf{W} = \mathbb{E}[W(t)] \). As the graph \( \mathcal{G}_W \) is strongly connected, the sequence \( W(t) \) solves the probabilistic average consensus problem, i.e., we have

\[ \mathbb{P}(\lim_{t \to \infty} x_k(t) = \alpha_k) = 1 \]

for all \( k \). (See Theorem 3.3 in [14].)

IV. SIMULATIONS

In this section, we report some simulation results to demonstrate the feasibility of the distributed fusion schemes outlined in Section III. A simple illustrative example will first be given in Section IV-A. A distributed classification application will then be described in Section IV-B.

A. Illustrative example

As an illustrative example, consider the strongly connected graph shown in Figure 1. The initial mass functions on a two-element frame \( \Omega = \{\omega_1, \omega_2\} \) are shown in Table I. The initial values \( x_{ik}(0) = \log Q_i(A_k) \) of the state variables, with \( A_1 = \{\omega_1\} \), \( A_2 = \{\omega_2\} \) and \( A_3 = \{\omega_1, \omega_2\} \), as well as the averages \( \alpha_k \) are given in Table II. The weight matrix (9) for the static consensus mechanism is

\[ W = \begin{pmatrix} 0.417 & 0.333 & 0.25 & 0.00 & 0.00 \\ 0.333 & 0.667 & 0.00 & 0.00 & 0.00 \\ 0.250 & 0.000 & 0.25 & 0.25 & 0.00 \\ 0.000 & 0.000 & 0.25 & 0.75 & 0.00 \\ 0.000 & 0.000 & 0.25 & 0.00 & 0.75 \end{pmatrix}. \]

Figure 2 shows the evolution as a function of time of state variables \( x_{ik} \) for \( i \in [1, 5] \) and \( k \in [1, 3] \). Each variable \( x_{ik}(t) \) converges to the mean value \( \alpha_k \) defined by (5). After convergence, each node can compute the combined commonality function (up to a multiplicative constant) as

\[ Q(\{\omega_1\}) = 0.111, \quad Q(\{\omega_2\}) = 2.65 \cdot 10^{-2}, \quad Q(\{\omega_1, \omega_2\}) = 1.24 \cdot 10^{-5}. \]

The unnormalized combined mass function computed from (4) is

\[ m^*(\{\omega_1\}) = 0.862, \quad m^*(\{\omega_2\}) = 0.0265 \quad \text{and} \quad m^*(\{\omega_1, \omega_2\}) = 1.24 \cdot 10^{-5}. \]

After normalization, we get

\[ m^*(\{\omega_1\}) = 0.808, \quad m^*(\{\omega_2\}) = 0.192 \quad \text{and} \quad m^*(\{\omega_1, \omega_2\}) = 0.008. \]

\[ \begin{array}{c|ccccc} \hline \{\omega_1\} & x_{11}(0) & x_{21}(0) & x_{31}(0) & x_{41}(0) & x_{51}(0) \\ \{\omega_2\} & x_{12}(0) & x_{22}(0) & x_{32}(0) & x_{42}(0) & x_{52}(0) \\ \{\omega_1, \omega_2\} & x_{13}(0) & x_{23}(0) & x_{33}(0) & x_{43}(0) & x_{53}(0) \\ \hline \end{array} \]

\[ \begin{array}{c|ccccc} \hline \{\omega_1\} & -0.124 & -0.136 & -0.998 & -0.466 & -0.473 \\ \{\omega_2\} & -1.882 & -0.097 & -0.435 & -0.554 & -0.663 \\ \{\omega_1, \omega_2\} & -3.325 & -0.248 & -4.150 & -1.599 & -1.977 \\ \hline \alpha_k & -0.124 & -0.136 & -0.998 & -0.466 & -0.473 \\ \{\omega_2\} & -1.882 & -0.097 & -0.435 & -0.554 & -0.663 \\ \{\omega_1, \omega_2\} & -3.325 & -0.248 & -4.150 & -1.599 & -1.977 \\ \end{array} \]
8.99 \cdot 10^{-5}$, which is the orthogonal sum of the initial mass functions $m_i$.

Figure 3 shows the evolution of state variables $x_{ik}(t)$ for the dynamic consensus scheme, where each link $(i,j)$ is picked at random at each time step, and the states of nodes $i$ and $j$ are updated using matrix (10). As we can see, the system converges to the same state as with the static consensus scheme.

**B. Distributed classification**

Let us now consider a simple distributed classification problem with two classes $\Omega = \{\omega_1, \omega_2\}$ and $n$ decision nodes. Each node $i$ collects a measurement $y_i$, which is a realization of a random variable $Y_i$ with class-conditional densities $f_{i1}(y)$ and $f_{i2}(y)$. If the prior class probabilities are unknown, then an unnormalized mass function on $\Omega$ given $Y_i = y_i$ can be computed using the Generalized Bayes Theorem (GBT) [17], [18], [19] as

$m_i(\emptyset) = (1 - cf_1(y_i))(1 - cf_2(y_i))$

$m_i(\{\omega_1\}) = (1 - cf_2(y_i))f_1(y_i)$

$m_i(\{\omega_2\}) = (1 - cf_1(y_i))f_2(y_i)$

$m_i(\Omega) = c^2 f_1(y_i)f_2(y_i)$,

where $c$ is a positive constant ensuring that $cf_1(y) \leq 1$ and $cf_2(y) \leq 1$ for all $y$. This mass function quantifies one’s beliefs in the class of the object given $y_i$ and without any prior belief on the class. Combining $m_i$ with a prior probability mass function $\pi$ on $\Omega$ yields the Bayesian posterior distribution, with

$(m_i \oplus \pi)(\{\omega_k\}) = p(\omega_k|y_i) \propto f_k(y_i)p(\omega_k), \quad k = 1, 2$.

To illustrate in this context the behavior of the consensus mechanisms outlined in Section III, we considered the communication graph with 10 decision nodes shown in Figure 4. We assumed $Y_i$ to have normal class-conditional distributions:

$Y_i|\omega_k \sim \mathcal{N}(\mu_k, \sigma_k), \quad k = 1, 2,$

with $\mu_1 = 0$, $\sigma_1 = 1$, $\mu_1 = 0.1$, $\sigma_1 = 1.1$. As an example, we considered the following measurement values:

0.56, 0.54, 1.01, 1.07, 0.29, 2.47, 0.49, 0.16, 0.64, −0.18

Figure 5 shows the masses $m_i(\{\omega_1\})$, $m_i(\{\omega_2\})$ and $m_i(\Omega)$ as a function of the number of iterations, for the static consensus mechanism. To compute a mass function for each node at each time step, we converted the state vectors $x_i$ into mass functions using the procedure outlined above.

Figure 6 displays the same information in a barycentric plot, where a mass function $m_i$ is represented as the barycenter of the three vertices of the triangle, with weights $m_i(\{\omega_1\})$. Fig. 2. Convergence of state variables $x_{ik}(t)$ to the averages $\alpha_k$ for the static consensus mechanism.

Fig. 3. Convergence of state variables $x_{ik}(t)$ to the averages $\alpha_k$ for the dynamic consensus mechanism.

Fig. 4. Communication graph for the distributed classification example.
Fig. 5. Convergence of masses $m_i(\{\omega_1\})$, $m_i(\{\omega_2\})$ and $m_i(\{\Omega\})$ to the orthogonal sum, for the static consensus mechanism.

$m_i(\{\omega_2\})$ and $m_i(\{\Omega\})$. The vacuous mass function $m_2$ such that $m_2(\{\Omega\}) = 1$ corresponds to the upper vertex of the triangle, while the certain mass functions focussed on $\{\omega_1\}$ and $\{\omega_2\}$ correspond, respectively, to the lower left and right vertex. The initial mass functions (shown as circles) converge to the mass function represented by a triangle, which is the orthogonal sum of the initial mass functions,

$$m^*(\{\omega_1\}) = 0.391, m^*(\{\omega_2\}) = 0.604,$$

$$m^*(\{\Omega\}) = 0.005.$$

A maximum-plausibility classifier would then select class $\omega_2$ based on the complete information. From Figure 5, we can see that the same decision would be reached by all the nodes after only 10 iterations, i.e., long before the network has actually converged to the belief consensus. As shown in Figure 7, a similar behavior is obtained using the asynchronous symmetric gossip mechanism.

V. CONCLUSION

We have shown that Dempster’s rule of combination can be implemented in a distributed way, using synchronous or asynchronous linear consensus mechanisms. This finding makes it possible to design distributed data fusion schemes based on Dempster-Shafer theory. The intended application of this theoretical work is collaborative perception in a fleet of intelligent vehicles [11]. Other potential applications include distributed decision in sensor networks [20], and modeling the dynamics of beliefs in social networks [21]. Taking into account the degrees of confidence of each of the agents in the other agents, and implementing alternative combination rules are open problems left for further research.

Fig. 6. Barycentric plot of the mass functions in the distributed classification example. Each mass function $m_i$ at a given iteration is represented as the barycenter of the three vertices of the triangle, with weights $m_i(\{\omega_1\})$, $m_i(\{\omega_2\})$ and $m_i(\{\Omega\})$. The initial mass functions are shown as circles, and the final one is shown as a triangle. The solid lines correspond to the trajectory of the mass functions from the initial state to the consensus state.

Fig. 7. Convergence of masses $m_i(\{\omega_1\})$, $m_i(\{\omega_2\})$ and $m_i(\{\Omega\})$ to the orthogonal sum, for the dynamic (symmetric gossip) consensus mechanism.
ACKNOWLEDGMENT

This research was supported by the Labex MS2T, which was funded by the French Government, through the program “Investments for the future” by the National Agency for Research (reference ANR-11-IDEX-0004-02).

REFERENCES